AD-A196 150

ROYAL SIGNALS AND RADAR ESTABLISHMENT

Memorandum 4134

TITLE: **CONNECTED DIGIT RECOGNITION IN A MULTILINGUAL ENVIRONMENT**

AUTHOR: R. K. Moore

DATE: February 1988

SUMMARY

The several languages and language dialects used by the member nations of NATO generate many questions regarding the use of speech recognition systems for NATO related applications. Two questions which immediately arise are: (a) are speech recognition systems sensitive to the differences between the several languages even though each system may be trained on a specific language of interest, and (b) if an operator of a speech recognition system uses a language other than his/her primary language, is the recognition performance of the system compromised by the use of the secondary language? Research Study Group 10 of NATO (AC/243, Panel III) addressed these and other questions by conducting a series of speech recognition tests using a multiple-language speech database. This report concentrates on the connected digit recognition performance.

The text of this memorandum is also published as Project II Final Report NATO AC/243 (Panel 3/RSG 10) D/11, 1987.

CONTENTS

1. INTRODUCTION

The several languages and language dialects used by the member nations of NATO generate many questions regarding the use of speech recognition systems for NATO related applications. Two questions which immediately arise are:-

a) are speech recognition systems sensitive to differences between the several languages even though each system may be trained on a specific language of interest, and

b) if an operator of a speech recognition system uses a language other than his/her primary language, is the recognition performance of the system compromised by the use of the secondary language?

Research Study Group 10 of NATO (AC/243, Panel III) addressed these and other questions by conducting a series of speech recognition tests using a multiple-language speech database. The database was provided to representatives of member nations who then arranged for it to be used on as many speech recognition systems as possible, given certain time constraints. Isolated digit utterances and connected digit sequences appeared in the database; results from the isolated word recognition systems have been reported elsewhere [1].

This report concentrates on the connected digit recognition performance.

2. SPOKEN DIGIT DATABASE

Nineteen talkers from five NATO countries (France, West Germany, the Netherlands, the United Kingdom and the United States) provided the recorded utterances. Fourteen of the speakers were male and five were female. Each speaker produced isolated and connected digit utterances in his/her primary language. Eleven speakers also produced the digit utterances in a secondary language. The secondary language was either English or French.

The connected sequences were spoken in groups of three, four and five digits, and speakers were prompted by reading from a set of pre-prepared lists. Table I indicates the amount of material that was obtained in this manner from each speaker. Apart from two isolated digit lists (which were intended to be used to train the recognisers), each list contained a randomised selection of tokens.

Since each recording session involved a considerable investment in time, it was agreed that some speakers should only record a reduced set of lists (see table I). The complete database thus contains eleven full sets and eighteen reduced sets of data.

Table II gives details of the speaker-language combinations contained in the database. It also shows which speakers spoke in a secondary language and which speakers spoke the full set of lists.

The RSG10 database thus contains a total of 18,000 utterances (37,300 digits). A complete description of the database has been reported elsewhere [2].

Group Size	No. of Lists	Groups/ List	Total No. of Groups	Total No. of Digits
1	5 (3)	100	500 (300)	500 (300)
3	4 (2)	50	200 (100)	600 (300)
4	2 (0)	50	100 (0)	400 (0)
5	2 (1)	50	100 (50)	500 (250)
		TOTALS:	900 (450)	2000 (850)

Table I: Amount of material provided by each speaker; the figures in brackets indicate the number of items in the reduced set.

Country	Speaker	Sex	Language	Full/ Reduced Set
US	KJ	M	E	F
"	SS	M	E	R
"	JP	M	E	R
"	MP	F	E	R
Neth.	LP	M	D/E	F
"	TV	M	D/E	R
"	LD	F	D/E	R
France	JG	M	F	R
"	JM	M	F/E	F
"	DT	M	F	R
"	FN	F	F/E	R
Germany	HO	F	G/E	R
"	GG	M	G/E	R(G)/F(E)
"	HK	M	G/E	R
"	BB	M	G/E	R
UK	MW	M	E/F	F
"	MT	M	E	F
"	GR	F	E	F
"	RM	M	E	F

Table II: Speakers and languages in the RSG10 spoken digit database (E-English, D-Dutch, F-French, G-German).

3. RECOGNITION SYSTEMS

The recognition systems employed in the study included four commercially available connected word recognition systems, a team of human listeners and a software-only system which implements a standard or 'reference' algorithm.

3.1 Laboratory 'Reference' System

The laboratory system used in the experiments employed the 'one-pass' connected word recognition algorithm based on 'dynamic time warping' [3]. The algorithm was implemented in software at the Royal Signals and Radar Establishment (RSRE) in the UK as part of a general laboratory facility for comparing different variants of word recognition algorithm. The system did not function in 'real-time'.

All of the speech material was digitised and annotated in advance of the recognition experiments. This meant that training, testing and scoring the results of the recogniser were able to be completely automated.

The recogniser itself was based on a 'textbook' version of the one-pass algorithm, and its performance can thus be regarded as providing a baseline which all other recognition schemes should be able to outperform.

3.2 Commercial Recognisers

3.2.1 NEC DP100

The Japanese DP100 was the first commercially available connected word recogniser. Its 1982 cost was about $60k. Based on the 'two-level' connected word recognition algorithm [4], it can be set up either to allow any number of words in a connected sequence or to only accept a specific number. It also has a facility for setting a reject threshold. The largest vocabulary the DP100 can accommodate is 120 words.

The DP100 was tested in the USA at the Rome Air Development Center (RADC) and, of the commercial recognisers, this is algorithmically the most similar to the 'reference' system described above.

3.2.2 MOZART RME88

The MOZART connected word recogniser is based on a 'dynamic time warping' algorithm developed at the Laboratoire d'Informatique pour la Mecanique et les Sciences de l'Ingenieur (LIMSI) in France [5]. The device is manufactured by VECSYS Ltd. and the publicly released version (RME186) cost $3.7k in 1982. Both versions have a reject facility and are able to operate in various modes in addition to connected word recognition; for example, word spotting and shadowing modes. The device also allows 'embedded' training (training on connected words). The vocabulary capability of the RME186 is 300 words.

The RME88 version of MOZART was tested in France at LIMSI.

3.2.3 ADES-III

ADES-III is a German connected word recogniser which is based on template matching using 'dynamic time warping'. Its maximum vocabulary is 230 words. It is different to the other recognisers in that it performs an explicit segmentation of words using statistical principles which are vocabulary dependent but speaker independent [6]. Like the DP100, ADES-III can operate with the phrase length being known (formatted) or unknown (unformatted) to the recogniser. Reference templates are constructed using an averaging process [7].

This recogniser was tested in Germany at AEG-Telefunken.

3.2.4 VERBEX V-1800

The V-1800 is a statistical recogniser based on the principles of 'hidden Markov modelling'. It cost about $86k in 1982. Like the other devices, the V-1800 has an adjustable reject threshold. This recogniser starts with a universal template for each word which is then updated by the repeated presentation of a set of speaker-specific training material. The optimal training time can be up to one minute per word. The system was tested in the USA at Verbex Co. by staff from RADC.

3.3 Human Listeners

As a control condition, part of the database was presented to human listeners in the quiet and in two levels of background noise (with signal-to-noise ratios of -3 and -9 dB). These experiments were conducted at the Institute for Perception TNO in the Netherlands and are described in detail elsewhere [8].

4. EXPERIMENTAL PROCEDURE

The recorded speech material was presented to each of the systems according to agreed guidelines. The guidelines specified the records to be taken during the experiments and contained instructions on how to run the experiments.

Two lists of isolated digits (per speaker-language combination) were reserved for training and tuning. Systems with a facility for 'embedded training' could optionally use a designated 3-digit list.

After the training phase, no parameter was to be adjusted unless an independent test made such a change necessary (for example, a level meter showing that the signal was too high or too low). It was stressed that such changes should not be made on the basis of recognition results.

In order to avoid a recogniser simply throwing out a test sample, the rejection threshold was to be set as low as possible.

Some systems (for example, the V-1800) could process even the fastest lists in real-time, but others (for example, MOZART) required occasional tape stops in order to allow the recogniser to catch up.

4.1 Training

4.1.1 Reference System

The laboratory 'reference' system used one reference template per spoken digit (for each speaker-language combination) and that template was formed from a single isolated utterance selected for its ability to discriminate between other isolated digit training utterances. This method has been shown to give better performance than using arbitrary single templates [9].

4.1.2 NEC DP100

Training for the DP100 was accomplished using one of the isolated digit training tables for each speaker-language combination. Although the DP100 nominally requires only one training example per digit, ten training passes were used (that is, there were ten templates per digit with no averaging).

4.1.3 MOZART RME88

The RME88 system was trained on isolated and connected digits. For isolated digit recognition two isolated digit templates were used for each word. For connected digit recognition one isolated digit template for each word was combined with up to 110 embedded words extracted from the three-digit training list in word spotting mode.

4.1.4 ADES-III

For ADES-III, two segmentation classifiers were generated from all of the connected training lists: one for male speakers and one for female speakers. A template for each word for each speaker was then constructed by averaging two examples taken from the appropriate three-digit training list.

4.1.5 VERBEX V-1800

The V-1800 training starts with a universal template for each word which is originally in English. In order to conduct tests in the other langages, samples were sent to Verbex Co. to enable them to construct a set of language-specific universal templates.

The device was then trained for each speaker-language combination using the two isolated digit training lists and the designated three-digit training list. Full lists were used. Interestingly, the V-1800 training algorithm performs recognition during the training phase.

First, an isolated digit training list was used to generate templates by comparison with the previously loaded universal templates. Then the system was trained using the second isolated digit training list and the training data was saved. If recognition errors occured during the training, the two lists were

played again (to a maximum of three passes). Training on the connected digit list used a minimum of two and a maximum of six passes.

The V-1800 system prompts the user to adjust the gain if necessary.

4.2 Testing

For each system, testing was performed on digit sets not used for training. Reject thresholds set as low as possible.

For both the DP100 and ADES-III, the number of digits to be input in each group was specified in advance ('formatted' input mode). Similarly, the MOZART system knew the maximum number of words to expect in a sequence.

The ADES-III system used 'continuous adaptation' and was thus effectively training on the test samples after they had been used for the test. Also, ADES-III was only tested on data from those speakers who spoke English (native and non-native).

The human listeners were tested on recordings by one male and one female from each country: ten speakers in English and two speakers in Dutch.

Neither the VERBEX V-1800 or the human listeners were tested on the 4-digit groups.

4.3 Scoring Method

Two types of error were regarded as important: group errors and digit errors. A group error occurs when a digit sequence contains one or more digit errors and this is easy to determine. For digit errors, real difficulties arise when there is more than one error in a group. For example, if the response to "382" is "321" then one interpretation is that two substitution errors have occured ("8" recognised as "2" and "2" recognised as "1"), but it is also possible that the "8" has been missed and the "1" inserted.

Two scoring methods were considered, one based on a strict comparison of the order of digit responses (thus "123" recognised as "23" would be interpreted as "1" recognised as "2, "2" recognised as "3" and "3" missed) and one based on string matching using dynamic programming. The first method tends to give a pessimistic assessment of the errors, whilst the second method tends to give an optimistic assessment.

In the absence of a more accurate scoring method, the agreed procedure was based on the first method described above since it could be applied automatically (without relying on human judgement), but was simple enough to be applied by hand (without a computer). However, it was decided to concentrate on group errors whenever possible, and to recommend that a computer program should be developed to analyse error patterns in a way that agrees with intuition and is useful in understanding the reasons for the errors.

4.4 Problems

Inevitably, with such a large experiment involving several laboratories in different countries, a considerable number of unforeseen events conspired to make the collation of results a difficult task. For example, speaker DT only recorded the first 14 groups of table 5A, the German tape broke during the US testing of the V-1800, and some machines have a problem if there is a pause within a group of digits.

Also, the fact that some speakers recorded more data than others and that some recognisers were only tested cn part of the database all contributed to the overall difficulties.

There was also an enormous am unt of material which had to be combined in order to construct the summaries presented in the next section. However, a great deal was learnt about the methodology of conducting large-scale speech recognition tests, and a report discussing this aspect of the project may be released at a later date.

5. RECOGNITION RESULTS

5.1 Overall Performance

The overall performance of the six different recognition systems is shown in figure 1. Percentage error rate is shown averaged across all speakers, all languages and all group sizes. From the figure it can be seen that the reference system has the most errors and, perhaps suprisingly, the best performance is not achieved by the human listeners but by the statistically-based recogniser: the VERBEX V-1800. Of the others, the rank ordering agrees with the degree of sophistication of the algorithms. As to the actual level of errors, even the best system has an overall error rate of just over 2% and this may be considered quite high for use in real applications.

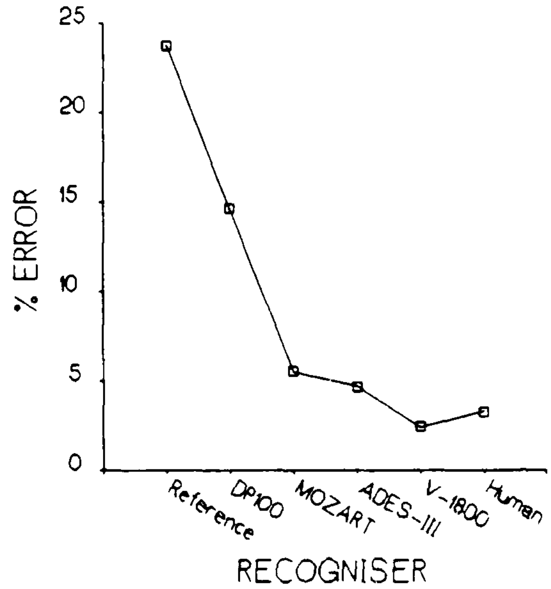

Figure 1: Overall performance.

Figure 2 illustrates the performance of the individual recognisers as a function of group size. Clearly the more digits there are in a group the more likely it is that there will be a recognition error. All of the systems show very good performance for isolated digits (group size = 1). However, the error rate for the reference system and the DP100 for connected digits is between 20 and 40% which is rather poor but neither of these systems used embedded training. Nevertheless, even the V-1800 makes 10% errors on five-digit groups.

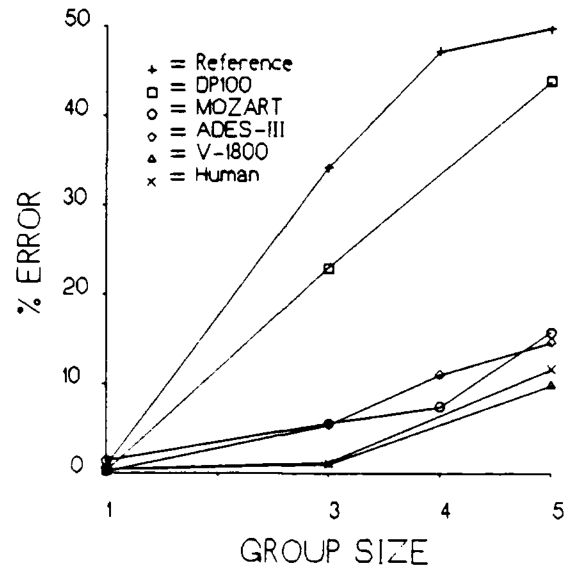

Figure 2: Effect of Group Size.

All of the recognisers agreed that the middle digit in a three-digit group is more difficult to recognise. However, the behaviour for four- and five-digit groups is difficult to assess because of the scoring method that was adopted (see section 4.3).

5.2 Effect of Male/Female

Figure 3 shows the results for the different recognisers analysed according to the sex of the speaker. Two systems recognise males better than females, two systems recognise females better than males and two others show no difference at all. The conclusion is thus that there is no significant difference in performance as a function of the sex of the speaker.

5.3 Effect of Language

The effect of the different languages used is depicted in figure 4. It can be clearly seen that there does seem to be a pattern to the overall behaviour as a function of language. A majority of the recognisers found the Dutch digits the most difficult to distinguish, with French the next most difficult and English the next. German digits were generally found to be the easiest.

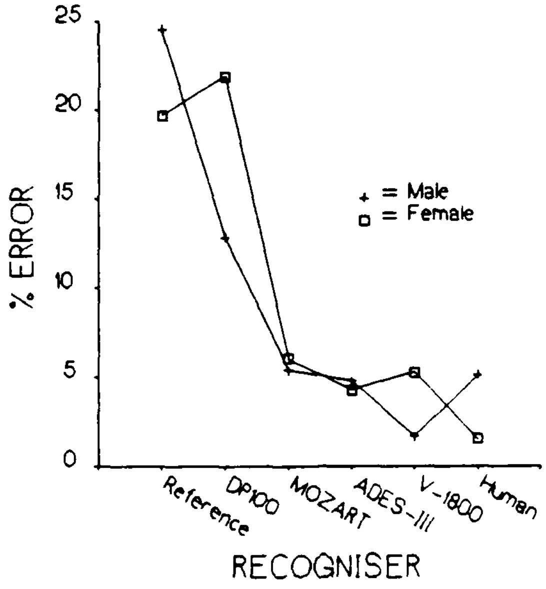

Figure 3: Effect of Sex.

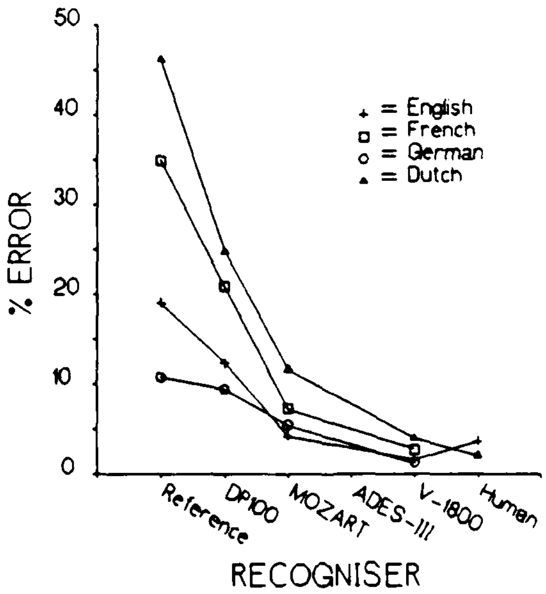

Figure 4: Effect of Language.

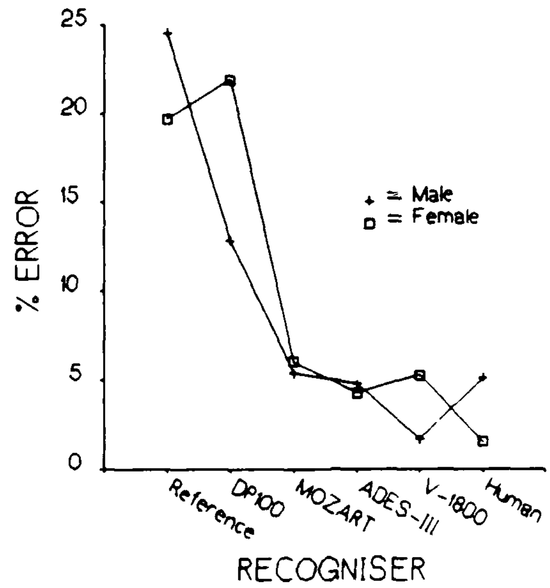

Figure 3: Effect of Sex.

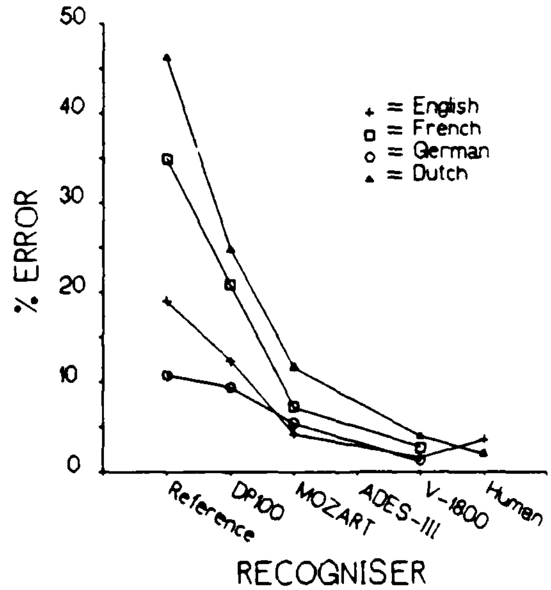

Figure 4: Effect of Language.

However, it is important to note that the spread of behaviour over different speakers (section 5.5) may account for most of these apparent language differences.

This latter point is partially confirmed by Human Equivalent Noise Ratio (HENR) analysis of the phonetic confusibility of the transcribed digits [10]. This analysis predicts that there is very little difference in confusability between the four languages; for example, a difference of one or two percent in error rate would only be detectable at an extremely poor signal-to-noise ratio such as -10dB. In such circumstances the HENR analysis predicts that Dutch would be easiest to recognise, German next, then French and that English would be the hardest - a prediction which is almost completely the reverse of the results presented in figure 4 for the automatic recognisers, but which is in agreement with the results obtained from the human listeners.

The behaviour of the connected word recognisers therefore depends more on the characteristics of the individual speakers than on the particular language that they used.

5.4 Effect of Native/Non-Native

For all of the recognisers, the recognition accuracy for connected digits spoken in a secondary language (mostly English) appears to be poorer than the recognition accuracy for those spoken in the primary language (see figure 5).

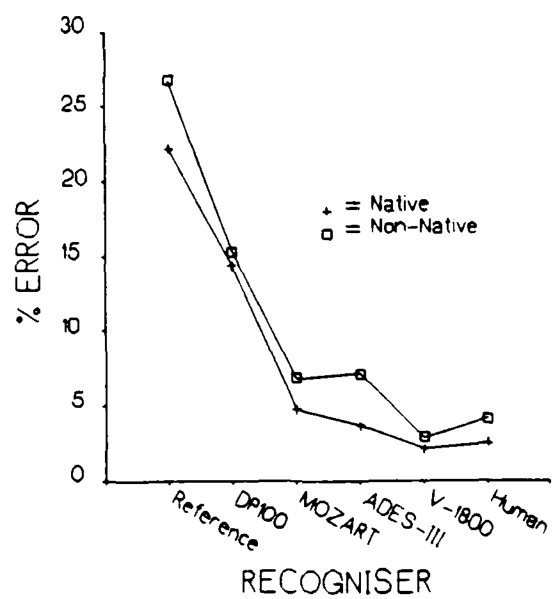

Figure 5: Effect of Native and Non-Native.

Unfortunately, this result is also heavily influenced by the characteristics of the individual speakers in the database. Close inspection of the results for those speakers who spoke both a primary and a secondary language reveals that there was a tendency for the recognisers to make fewer errors on the secondary language! This result is confirmed by a consistency analysis [11] of these particular speakers. The behaviour apparent in figure 5 must therefore be an artifact arising from the comparatively low error rates being achieved by those speakers who only spoke in their native language.

The conclusion must therefore be that most speakers are more consistent in their secondary language than in their primary language, probably due to an increase in effort being applied to the less familiar task.

5.5 Effect of Speaker

Inspection of figure 6 indicates that variable performance results both within and between languages and within and between recogisers. For example, speaker SS from the United States had excellent results on all recognisers, while speakers HO and MW were recognised rather poorly. The overall speaker ranking agrees quite well with that obtained using consistency analysis [11].

Also, individual speakers exhibited divergent results: for example, speakers MP and GR have very good performance on some recognisers and very poor performance on others.

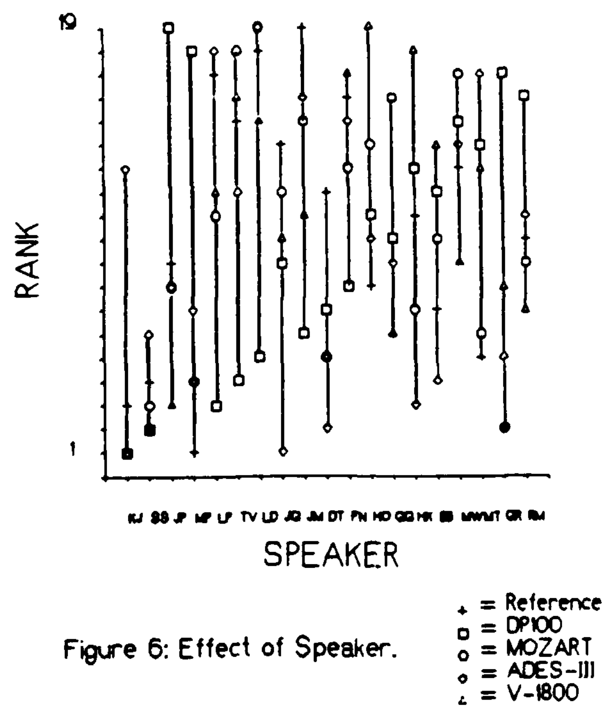

Figure 6: Effect of Speaker.

Some of the factors which influenced an individual speaker's results were:-

a) some speakers spoke rather rapidly (in particular JM, LP and HO),

b) other speakers spoke rather slowly (in particular GR),

c) occasional pauses were inserted within an utterance sequence (causing some recognisers to split the sequence and miss the last part),

d) the last digit in a sequence was often rather low in level.

6. DISCUSSION

It is very difficult to say anything about the significance of the results. That is, it is not clear whether the results are sufficiently general to predict the results of similar tests using a different database. The number of test utterances, on its own, is not an indication of significance if only a small - possibly unrepresentative - training set was used. If one regards the recognition process as matching one set of utterances with another, it can be argued that as well as the training and test utterance sets being large, they should also be approximately equal in size for a statistically balanced experiment. To make use of large training sets, the experiments would have to be run repeatedly using different subsets of the training data until all the tokens were used up. This would create a large experiment and automatic data processing and handling would become essential.

At best the tests carried out so far enable something to be said about isolated versus connected word recognition, about different languages and speakers, and about the recognition of digits in general. The database was established in a quiescent environment without interaction between the speaker and a system. The results tell us nothing about the effect of miltary environments on speech, or their direct effect on the speech signal, or how these will affect recognition performance. It remains for future experiments to predict likely recognition performance for military applications.

There are many difficulties in comparing performance of different machines, both generally and in the particular case of the results reported here. The difficulties arise from several sources:-

a) Statistical problems - are the observed differences in error rate sufficient to indicate real differences in recognition capability?

b) Mandatory differences in procedure dictated by the speech recognition systems, for example: single or multiple example utterances per template.

c) Differences in facilities that were exploited by the experimenters, for example: use of multiple templates.

d) Arbitrary differences in procedure.

The biggest difficulty in comparing performance is that different subsets of the training data may be used by different machines. Again this problem is

exacerbated by small training sets, increasing the likelihood that a particular set used by a particular machine will be unrepresentative of the general case.

Scoring the recognition errors appeared to be the most troublesome aspect of the testing exercise. Once a group error was identified, classifying an intra-group error type according to the method outlined in section 4.3 often produced interpretations which disagreed with intuition. For example, if the response to "59437" was "9437", the agreed method classified this as four digit substitution errors ("5" as "9", "9" as "4", "4" as "3" and "3" as "7") and one missed "7". As a consequence, it was not considered appropriate to study individual digit errors as a function of their position in a digit sequence - something which would otherwise have been of considerable interest.

Recent results obtained in the UK using a MARCONI SR128 connected word recogiser indicate that the pattern of errors is not consistent on repeated runs of the same conditions; the error rate for a table can vary by a factor of two. At present it must be assumed that this behaviour is typical of the other systems, so any difference in error rates under different conditions that are within a factor of two must be interpreted with some care.

It is also clear that all of the factors such as male/female, native/non-native, English/French/German/Dutch are swamped by speaker variability. So again care must be exercised in the interpretation of the results.

Finally, the database of connected digit sequences does seem to have been a good test corpus; it presented a reasonable level of difficulty and even such a modest vocabulary gave rise to a mammoth assessment task (underlining the importance of developing automated test procedures).

7. RECOMMENDATIONS

a) Design future experiments to give statistically significant results.

b) Design future experiments specifically for the set of machines being compared so that differences in training and operating strategies can be considered from the outset.

c) Use automatic data handling and processing to reduce the effort required for analysing an experiment.

d) The manner in which subjects speak should be categorised (for example, loud versus soft, slow versus fast). This may result in valuable guidelines on how to train speakers to obtain maximum recognition performance.

e) Introduce environmental factors such as noise, vibration and 'g'.

8. CONCLUSIONS

Connected digit recognition is much more difficult than isolated digit recognition.

One system performed better than the others, but if its performance on this data was typical of real applications then it might not be good enough. However, it must NOT be assumed that the performance achieved on this data will be repeated in real applications - real-life performance might be better or worse. This is because current recognisers do not allow for (a) the adverse effects of the environment on a user's manner of speaking, (b) directly added noise, (c) the ability of a user to compensate for the environmental effects or (d) a user's ability to learn, over a period of time, how to speak to obtain the best recognition performance from a particular machine.

All the results show that extensive training of a system gives good results whichever language is used, and even if the speaker is speaking a secondary language.

In the future speech recognisers may exist that do not merely match data sets but have preconceived ideas about speech, so that the training template set is not wholly a function of the training utterances. This leads ultimately to speaker independent systems where no training utterances are required at all. There will also be adaptive systems, particularly useful for military environments where it might be difficult to obtain training tokens in an operational situation. The recognition performance of adaptive machines can thus be expected to improve as recognition tests proceed.

All these categories of machine pose new questions about defining and measuring recognition performance so that comparisons can be made with conventional speech recognisers. The tests carried out so far have illustrated the difficulty of designing a generalised experiment. The variabilities of operating procedures across different machines always cast doubt on the validity of comparing results. It may be wiser to design future experiments specifically for the set of machines being compared so that differences in training routines and other characteristics can be considered from the outset.

In general, the data indicate more variance based on speaker differences than on any other differences, and that the better the training the better the recognition will be.

Annex I: REFERENCES

[1] Bridle J S. 'Connected word recognition for use in military systems', NATO AC/243-D/854; AC243 (Panel III) D/128, 1983.

[2] Vonusa R S, Nelson J T, Smith S E and Parker J G. 'NATO AC/243 (Panel III RSG10) Language Database', Proc. US NBS Workshop on Standardisation for Speech I/O Technology, pp 225-228, Gaithersburg, 1982.

[3] Bridle J S, Brown M D and Chamberlain R M. 'Continuous connected word recognition using whole-word templates', The Radio and Electronic Engineer, Vol.53, pp 167-175, 1983.

[4] Sakoe H. 'Two-level DP matching - a dynamic programming based pattern matching algorithm for connected word recognition', Trans. IEEE ASSP, Vol.27, pp 588-595, 1979.

[5] Gauvain J L, Lienard J S and Mariani J J. 'Un systeme de reconnaissance de mots enchaines et de detection mots dans la parole continue', Registered technical Report Anvar No.50505, April, 1982.

[6] Zelinski R and Class F. 'A segmentation technique of word groups by statistical principles', IEEE Trans. ASSP-31, No.4, pp 818-827, 1983.

[7] Zelinski R and Class F. 'A learning procedure for speaker dependent word recognition systems based on sequential processing of input tokens', Proc. IEEE Int. Conf. ASSP, pp 1053-1056, Boston, 1983.

[8] Pols L C W. 'How humans perform on a connected-digits database', Proc. IEEE Int. Conf. ASSP, pp 867-870, Paris, 1982.

[9] Russell M J, Deacon J C A and Moore R K. 'Some implications of the effect of template choice on the performance of an automatic speech recogniser', Proc. Inst. Acoustics Autumn Conf., Windermere, pp 287-292, November, 1984.

[10] Moore R K. 'Evaluating speech recognisers', Trans. IEEE ASSP, Vol.25, pp 178-183, 1977.

[11] Smith D C, Russell M J and Tomlinson M J. 'Rank-ordering of subjects involved in the evaluation of automatic speech recognisers', RSRE Memorandum No.3926, January 1986.

Annex II: LIST OF PARTICIPANTS

Speech Research Unit
Royal Signals and Radar Establishment
St. Andrews Road
Great Malvern
Worcestershire
WR14 3PS
United Kingdom

LIMSI-CNRS
BP30
91406 Orsay Cedex
France

AEG-Telefunken
Research Institute
Sedanstr. 10
D7900 Ulm
West Germany

Institute for Perception TNO
PO 23
Soesterberg 3769 DE
Netherlands

Rome Air Development Center - IRAA
Griffiss AFB
NY 13441
United States of America

DOCUMENT CONTROL SHEET

Overall security classification of sheet .... UNCLASSIFIED ....

(As far as possible this sheet should contain only unclassified information. If it is necessary to enter classified information, the box concerned must be marked to indicate the classification eg (R) (C) or (S))

1. DRIC Reference (if known)	2. Originator's Reference MEMO 4134	3. Agency Reference	4. Report Security Classification U/C
5. Originator's Code (if known) 778400	6. Originator (Corporate Author) Name and Location RSRE St Andrews Road, Malvern, Worcs. WR14 3PS		
5a. Sponsoring Agency's Code (if known)	6a. Sponsoring Agency (Contract Authority) Name and Location		

7. Title
CONNECTED DIGIT RECOGNITION IN A MULTI-LINGUAL ENVIRONMENT

7a. Title in Foreign Language (in the case of translations)

7b. Presented at (for conference papers) Title, place and date of conference

8. Author 1 Surname, initials	9(a) Author 2	9(b) Authors 3,4...	10. Date	pp. ref.
MOOR, R.K.			1988.2	17

11. Contract Number	12. Period	13. Project	14. Other Reference

15. Distribution statement

Descriptors (or keywords)

continue on separate piece of paper

Abstract The several languages and language dialects used by the member nations of NATO generate many questions regarding the use of speech recognition systems for NATO related applications. Two questions which immediately arise are: (a) are speech recognition systems sensitive to the differences between the several languages even though each system may be trained on a specific language of interest, and (b) if an operator of a speech recognition system uses a language other than his/her primary language, is the recognition performance of the system compromised by the use of the secondary language? Research Study Group 10 of NATO (AC/243, Panel III) addressed these and other questions by conducting a series of speech recognition tests using a multiple-language speech database. This report concentrates on the connected digit recognition performance.

The text of this memorandum is also published as Project II Final Report NATO AC/243 (Panel 3/RSG 10) D/aa. 1987.

S80/48